DOCUMENT RESUME

ED 398 248 TM 025 188

AUTHOR Thompson, Bruce
TITLE Problems with Multivariate Normality: Can the Multivariate Bootstrap Help?
PUB DATE 4 Apr 96
NOTE 48p.; Paper presented at the Annual Meeting of the Society for Applied Multivariate Research (Houston, TX, April 1996).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Computer Assisted Testing; Computer Software; *Correlation; *Statistical Distributions; *Test Use
IDENTIFIERS *Bootstrap Methods; *Multivariate Normal Distribution

ABSTRACT

Multivariate normality is required for some statistical tests. This paper explores the implications of violating the assumption of multivariate normality and illustrates a graphical procedure for evaluating multivariate normality. The logic for using the multivariate bootstrap is presented. The multivariate bootstrap can be used when distribution assumptions are not met, or for descriptive purposes in all cases. Multivariate bootstrap logic is illustrated for the canonical correlation case. From a practical point of view, computer automation is required for the bootstrap approach. Various types of software for conducting multivariate bootstrap analyses are described. An appendix presents computer program output for the data. (Contains 1 figure, 4 tables, and 29 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

swpa96.wpl 3/8/96

ED 398 248

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
BRUCE THOMPSON

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Problems with Multivariate Normality:

Can the Multivariate Bootstrap Help?

Bruce Thompson

Texas A&M University
and
Baylor College of Medicine

BEST COPY AVAILABLE

Paper presented at the annual meeting of the Society for Applied Multivariate Research (session #164), Houston, April 4, 1996.

TM025188

ABSTRACT

Multivariate normality is required for some statistical tests. A graphical procedure for evaluating multivariate normality is illustrated. The logic for using the multivariate bootstrap is presented. The multivariate bootstrap can be used when distribution assumptions are not met, or for descriptive purposes in all cases. The multivariate bootstrap logic is illustrated for the canonical correlation case. Various software for conducting multivariate bootstrap analyses is described and cited.

Researchers have increasingly recognized that multivariate analyses are vital in the social sciences, for at least two reasons (Fish, 1988; Thompson, 1994c). First, use of multivariate methods avoids the inflation in experimentwise Type I error rates that occurs when univariate methods are employed in a single study to test multiple hypotheses that are at least partially uncorrelated. Second, and more importantly, multivariate methods can be employed to analytically honor a substantive reality in which most effects have multiple causes, and most effects have multiple consequences.

For these reasons, multivariate methods are being employed with increasing frequency. For example, Emmons, Stallings and Layne (1990) studied 16 years of research reports in three journals, and found that

> the multivariate characteristic of the social science research environment with its many confounding or intervening variables has been addressed through the trend toward increased use of multivariate analysis of variance and covariance, multiple regression, and multiple correlation. (p. 14)

Similarly, Grimm and Yarnold (1995) recently noted that, "In the last 20 years, the use of multivariate statistics has become commonplace. Indeed, it is difficult to find empirically based articles that do not use one or another multivariate analysis" (p. vii).

The purpose of the present paper is to explore the

implications of violating the assumption of multivariate normality that is required in some multivariate applications. Marascuilo and Levin (1983) nicely summed up several important features of the assumption:

> The multivariate normal distribution is somewhat hidden throughout multivariate methods. It is not required in the estimation and data description aspects of the theory. Its impact and role, however, are basic to the [statistical significance] inference procedures of multivariate analysis and it is here that it must be assumed. There are no satisfactory tests of its truth in any one situation.

Although multivariate normality is not required to estimate most multivariate parameters (e.g., function coefficients, structure coefficients), even in these cases the distributions of the variables must be reasonably comparable. Multivariate parameters are estimated using the correlation or the variance/covariance matrix from the sample. As Thompson (1984, p. 17) noted, therefore even aside from the required assumptions for statistical significance testing,

> ...the magnitudes of the coefficients of the correlation [or covariance] matrix must not be attenuated by large differences in the shapes of the distributions for the variables. It is important to emphasize that... [parameter estimation usually]

> does not require that the variables be normally distributed as long as there is no substantial attenuation associated with distribution differences, regardless of what these distributions may be.

The present paper first reviews one method for estimating multivariate normality. Next, the use of the bootstrap in such situations is then explored. Finally, a heuristic example is presented.

Evaluating Multivariate Normality

It is important to note initially that evaluations of univariate and bivariate normality are not sufficient to establish that the assumption of multivariate normality has been met. As Thompson (1984, p. 18) noted,

> ...examining the univariate or the bivariate distributions will not conclusively resolve this uncertainty. Multivariate distributions can be nonnormal even when all subsets of univariate or bivariate distributions are normal, just as a bivariate distribution may be nonnormal even when both the individual variables are distributed in a normal matter.

One method of exploring whether the assumption of multivariate normality has been met involves a graphical procedure explained by Stevens (1986, pp. 207-212). Thompson (1990) provides a computer program that automates this procedure. Table 1 presents data used

to illustrate this approach. The heuristic example involves scores of 26 people on three variables.

INSERT TABLE 1 ABOUT HERE

Table 2 presents descriptive statistics for these data. Table 3 presents the Mahalanobis distances (D^2) of each of the score vectors for the 26 cases from the centroid, i.e., the Cartesian coordinate on each of the three variables associated with the means (6.40, 6.87, and 6.72, respectively). For example, case #8 had scores of 6.7, 6.0, and 7.2, which are close to the three means, respectively, thus resulting in the smallest D^2 value (0.694) for this case.

INSERT TABLES 2 AND 3 ABOUT HERE

In the graphical procedure these distances are sorted and associated chi-square and p values are computed, as illustrated in Table 4. Finally, the 26 pairs of chi-square and D^2 values are plotted, as illustrated in Figure 1. If a reasonably straight lines is defined within the plot, the data are taken to be multivariate normal.

INSERT FIGURE 1 ABOUT HERE

The Multivariate Bootstrap

Uses of the Bootstrap

As has been noted elsewhere,

> The bootstrap can actually be used in two somewhat discrete ways. First, it can be used *descriptively* to evaluate whether results are reasonably stable over different configurations of subjects... Second, the bootstrap can be used *inferentially*, if we consult all four statistics from our resampling analyses, and use them to empirically construct study-specific test distributions or confidence intervals. This is only another approach to testing statistical significance. (Thompson, 1993, pp. 372-373)

That is, when the assumption of multivariate normality cannot be met such that statistical significance testing cannot reasonably be conducted, one can use bootstrap methods to develop *study-specific* sampling distributions that can then be used in statistical tests.

Of course, statistical significance tests are of extremely limited utility (Carver, 1978; Cohen, 1994; Thompson, 1993, 1994a, 1996). These tests do <u>not</u> evaluate (a) the value or (b) the replicability of our results. These inferential tests do <u>not</u> evaluate either the probability of population parameters or the probability of the statistics in future samples!

But the bootstrap may still be valuable, even if we do not wish to perform statistical tests. The bootstrap may be used to explore the stability of parameter estimates when distributional problems may have attenuated certain relationships and consequently impacted the parameters estimated from these correlations or

covariances. Of course, the bootstrap may be useful in describing the replicability of results even when distributional assumptions are fully met.

Logic of the Bootstrap

The logic of the bootstrap has primarily been elaborated by Efron and his colleagues. Diaconis and Efron (1983) and Thompson (1994b) provide fairly accessible explanations. To make the present discussion concrete, let's presume that we had a sample of 50 subjects' scores on two variables, and that we wanted to estimate the Pearson r between the two variables. We would initially compute this statistic for the sample.

We would then draw a so-called random "resample" of 50 subjects from our original sample. But the trick is that we draw the resample *with replacement*. This means that our first subject may not be drawn at all in this resample. But subject #2 might be drawn several times. Thus, the resample consists of a different configuration of 50 subjects (some used multiple times) than our original sample. We would then compute the r in our resample.

When we randomly draw our resamples, we randomly select *all* the scores of each given resampled subject (i.e., in this case pairs of scores). And the reason that we draw *exactly* 50 subjects (at least in what should be at least one of the resampling strategies that we use) is to honor the influences of sampling error involved with sampling exactly 50 subjects.

Of course, what we can did once we could do a second time, by drawing a completely independent second random resample of 50

subjects' pairs of scores on the two variables. This would represent yet another configuration of the 50 original subjects' scores. We would then compute a second estimate of $\underline{r}$.

Over all of our resamples we could create a distribution of our estimates of $\underline{r}$. This would be an *empirically estimated* sampling distribution, rather than the theoretically assumed sampling distribution employed in conventional statistical tests (Arnold, 1996). And the standard deviation of the various resample parameter estimates is nothing less than an *empirically estimated* standard error of the statistic (i.e., SE_r). The ratio of the $\underline{r}$ in our original sample to this SE_r behaves like a $\underline{t}$ statistic. However, another alternative is to employ the sampling distribution to compute a confidence interval about our estimate.

It is conventional practice to resample at least 1,000 times, and 1,500 or 2,000 resamples would not be uncommon. More samples are especially important if our purpose is *inferential* (i.e., statistical significance testing), because here the tails of the sampling distribution are the focus, and considerably more subjects are required to adequately estimate these parts of the sampling distribution. Alternatively, in the *descriptive* application, our focus is basically on the question, "if we mix up our subjects in a whole lot of ways, do we still get basically the same estimate, no matter what we do?".

Obviously, from a practical point of view the bootstrap approach requires computer automation. Lunneborg (1987) has offered some excellent microcomputer programs that automate this logic for

univariate applications. In fact, user-friendly PC bootstrap software has become available from publishers around the world. Examples of such software and the distributors of the software include: (a) "Resampling Stats", distributed by Resampling Stats, 612 N. Jackson, Arlington, VA 22201; (b) "Statistical Calculator", distributed by Erlbaum, 27 Palmeira Mansions, Church Road, Hove East Sussex BN3 2FA, United Kingdom; (c) SPIDA, distributed on behalf of its Australian author by SERC, 1107 NE 45th--Suite 520, Seattle, WA 98105; and (d) the menu-driven program, BOJA, distributed by iecProGAMMA, P.O. Box 841, 9700 AV Groningen, The Netherlands.

A Multivariate Logic

The use of the bootstrap in univariate applications is quite straightforward. However, in multivariate analyses which produce multiple sets of estimates (e.g., two discriminant functions, three factors), special problems arise. In one resample a given factor may appear as the first function [equation, or factor], but in another resample may arise as the same construct but as the second function [equation, or factor]. Such variations are usually not substantively relevant or troubling, as long as the underlying constructs are invariant, but do create analytic problems.

In bootstrap applications using structural equation modeling, the solution is quite straightforward: simply use the matrix declaring the fixed and freed parameters to define a common factor space across all resamples. But in classical multivariate methods a solution is not as obvious.

Thompson (1995) explained the problem and proposed a solution:

> The major barrier to conducting a multivariate bootstrap involves the multidimensional character of the "space" in which the analysis is conducted. The bootstrap must be applied such that each of the hundreds or thousands of resampling results are all located in a common factor space before the mean, SD, skewness and kurtosis are computed.... If the analyst computed mean structure (or pattern) coefficients for the first variable on the first component across all the repeated samplings, the mean would be a nonsensical mess representing an average of some apples, some oranges, and perhaps some kiwi. The sampled solutions must be rotated to best fit positions with a common target solution, prior to computing means and other statistics across the [re]samples, so that the results are reasonable. (pp. 88-89)

In short, a "target" matrix is used to define a common factor space, and all resample results are rotated to best-fit position with this factor space using Procrustean rotation. Such applications can be generalized across classical parametric analyses, because all such analyses are special cases of canonical correlation analysis (Fan, 1992; Knapp, 1978; Thompson, 1984, 1991).

Thompson (1988, 1992, 1995) provides software for multivariate

applications (factor analysis, descriptive discriminant analysis, and canonical correlation analysis, respectively) that all invoke this solution. Borrello and Thompson (1989) and Scott, Thompson, and Sexton (1989) are examples of applications of the multivariate bootstrap.

Heuristic Example of the Multivariate Bootstrap

An application of the canonical bootstrap program, CANSTRAP (Thompson, 1995), is presented here as an illustration of the procedure. The illustration employs scores on six variables (i.e., four in one set, and two in the other set) from 50 cases from the Holzinger and Swineford (1939, pp. 81-91) data. These scores on ability batteries have classically been used as examples in both popular textbooks (Gorsuch, 1983, *passim*) and computer program manuals (Jöreskog & Sörbom, 1989, pp. 97-104), and thus are familiar to many readers.

Appendix A presents the program output for the data. First, canonical results are derived for the original sample. As reported in the appendix (p. 22 of the present paper), the 6x6 correlation matrix is computed, and partitioned into the quadrants associated with the variable sets. Also as reported in the appendix ("matrix to be analyzed", p. 23), the so-called 2x2 "quadruple-product" matrix is then computed (cf. Thompson, 1984).

The eigenvalues from the principal components analysis of the quadruple-product matrix are the two squared canonical correlation coefficients for these data (see p. 23). The two components are then used to compute canonical function coefficients (p. 23), and

subsequently canonical structure coefficients (p. 24) (Thompson & Borrello, 1985). The function coefficient matrix becomes the target matrix (p. 24) for Procrustean rotations for all the resamples.

Appendix A (pp. 24-30) presents full results for both the first two resamples. As noted previously, in each resample a given subject may be drawn not at all, or once, or multiple times. For example, in resample #1 (pp. 24-27) person 18 was drawn twice (as the 1st person in the resample and as the 14th person in the resample).

In the present example, 1,000 resamples were drawn. CANSTRAP then presents a description of the resampling process, so that randomness can be confirmed. For example, it is noted (p. 31) that person 1 was drawn once in resample #1, once in resample #2, not at all in resample #3, and twice in resample #991. Across the 1,000 resamples, person 1 was resampled 1,008 times (p. 31). The fewest times a person was resampled was 942; the most times was 1,056 (p. 32).

Then the multivariate bootstrap results are presented. The mean R_C^2 on Function I was .43958 (p. 32). The *empirically estimated* standard error of this statistic was .11744 (p. 32). Of course, when standard errors are empirically estimated, the standard errors may differ for different parameter estimates even when two parameter estimates are identical and sample size is a given fixed value.

The appendix also presents the mean function and structure coefficients (pp. 32-34) for this analysis. For example, across

1,000 resamples the structure coefficients of variables 3 and 4 on Function I were roughly equal (+.5515 and +.5407, respectively), while the standard errors for these estimates (.6003 and .4675, respectively) were not as equal. In this case the smaller of these two parameter estimates was somewhat more stable across the various 1,000 configurations of the original 50 subjects.

Summary

Multivariate normality is required for some statistical tests. A graphical procedure for evaluating multivariate normality was presented. The logic for using the multivariate bootstrap was presented. The multivariate bootstrap can be used when distribution assumptions are not met, or for descriptive purposes in all cases. The multivariate bootstrap logic was illustrated for the canonical correlation case. Various software for conducting multivariate bootstrap analyses was cited.

References

Arnold, M.A. (1996, January). The effects of two types of sampling error on common statistical analyses. Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED forthcoming)

Borrello, G., & Thompson, B. (1989). A replication "bootstrap" analysis of the structure underlying perceptions of stereotypic love. Journal of General Psychology, 116, 317-327.

Carver, R.P. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Cohen, J. (1994). The earth is round ($p < .05$). American Psychologist, 49, 997-1003.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. Scientific American, 248(5), 116-130.

Emmons, N.J., Stallings, W.M., & Layne, B.H. (1990, April). Statistical methods used in *American Educational Research Journal*, *Journal of Educational Psychology*, and *Sociology of Education* from 1972 through 1987. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA. (ERIC Document Reproduction Service No. ED 319 797)

Fan, X. (1992, April). Canonical correlation analysis as a general data-analytic model. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 348 383)

Fish, L. (1988). Why multivariate methods are usually vital.

Measurement and Evaluation in Counseling and Development, 21, 130-137.

Gorsuch, R.L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Erlbaum.

Grimm, L.G., & Yarnold, P.R. (Eds.). (1995). Reading and understanding multivariate statistics. Washington, DC: American Psychological Association.

Holzinger, K.J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution (No. 48). Chicago, IL: University of Chicago.

Jöreskog, K.G., & Sörbom, D./SPSS. (1989). LISREL 7: A guide to the program and applications (2nd ed.). Chicago: SPSS.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Lunneborg, C.E. (1987). Bootstrap applications for the behavioral sciences. Seattle: University of Washington.

Marascuilo, L.A., & Levin, J.R. (1983). Multivariate statistics in the social sciences: A researcher's guide. Monterey, CA: Brooks/Cole.

Scott, R.L., Thompson, B., & Sexton, D. (1989). Structure of a short form of the Questionnaire on Resources and Stress: A bootstrap factor analysis. Educational and Psychological Measurement, 49, 409-419.

Stevens, J. (1986). Applied multivariate statistics for the social sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Thousand Oaks, CA: SAGE.

Thompson, B. (1988). Program FACSTRAP: A program that computes bootstrap estimates of factor structure. Educational and Psychological Measurement, 48, 681-686.

Thompson, B. (1990). MULTINOR: A FORTRAN program that assists in evaluating multivariate normality. Educational and Psychological Measurement, 50, 845-848.

Thompson, B. (1991). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development, 24(2), 80-95.

Thompson, B. (1992). DISCSTRA: A computer program that computes bootstrap resampling estimates of descriptive discriminant analysis function and structure coefficients and group centroids. Educational and Psychological Measurement, 52, 905-911.

Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. Journal of Experimental Education, 61(4), 361-377.

Thompson, B. (1994a). The concept of statistical significance testing (An ERIC/AE Clearinghouse Digest #EDO-TM-94-1). Measurement Update, 4(1), 5-6. (ERIC Document Reproduction Service No. ED 366 654)

Thompson, B. (1994b). The pivotal role of replication in psychological research: Empirically evaluating the replicability of sample results. Journal of Personality, 62(2),

157-176.

Thompson, B. (1994c, February). Why multivariate methods are usually vital in research: Some basic concepts. Paper presented as a Featured Speaker at the biennial meeting of the Southwestern Society for Research in Human Development, Austin, TX. (ERIC Document Reproduction Service No. ED 367 687)

Thompson, B. (1995). Exploring the replicability of a study's results: Bootstrap statistics for the multivariate case. Educational and Psychological Measurement, 55, 84-94.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25(2), 26-30.

Thompson, B., & Borrello, G.M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Table 1
Data (n=26) From Stevens (1986, p. 209) Example

1	5.80000	9.70000	8.90000
2	10.60000	10.90000	11.00000
3	8.60000	7.20000	8.70000
4	4.80000	4.60000	6.20000
5	8.30000	10.60000	7.80000
6	4.60000	3.30000	4.70000
7	4.80000	3.70000	6.40000
8	6.70000	6.00000	7.20000
9	7.10000	8.40000	8.40000
10	6.20000	3.00000	4.30000
11	4.20000	5.30000	4.20000
12	6.90000	9.70000	7.20000
13	5.60000	4.10000	4.30000
14	4.80000	3.80000	5.30000
15	2.90000	3.70000	4.20000
16	6.10000	7.10000	8.10000
17	12.50000	11.20000	8.90000
18	5.20000	9.30000	6.20000
19	5.70000	10.30000	5.50000
20	6.00000	5.70000	5.40000
21	5.20000	7.70000	6.90000
22	7.20000	5.80000	6.70000
23	8.10000	7.10000	8.10000
24	3.30000	3.00000	4.90000
25	7.60000	7.70000	6.20000
26	7.70000	9.70000	8.90000

Table 2
Descriptive Statistics for the Table 1 Data

Means			
	6.40385	6.86923	6.71538
Variance/Covariance Matrix			
1	4.52279	3.98212	2.94114
2	3.98212	7.41261	3.70049
3	2.94114	3.70049	3.31015
Inverted Variance/Covariance Matrix			
1	0.56740	-0.12024	-0.36973
2	-0.12024	0.33075	-0.26292
3	-0.36973	-0.26292	0.92454

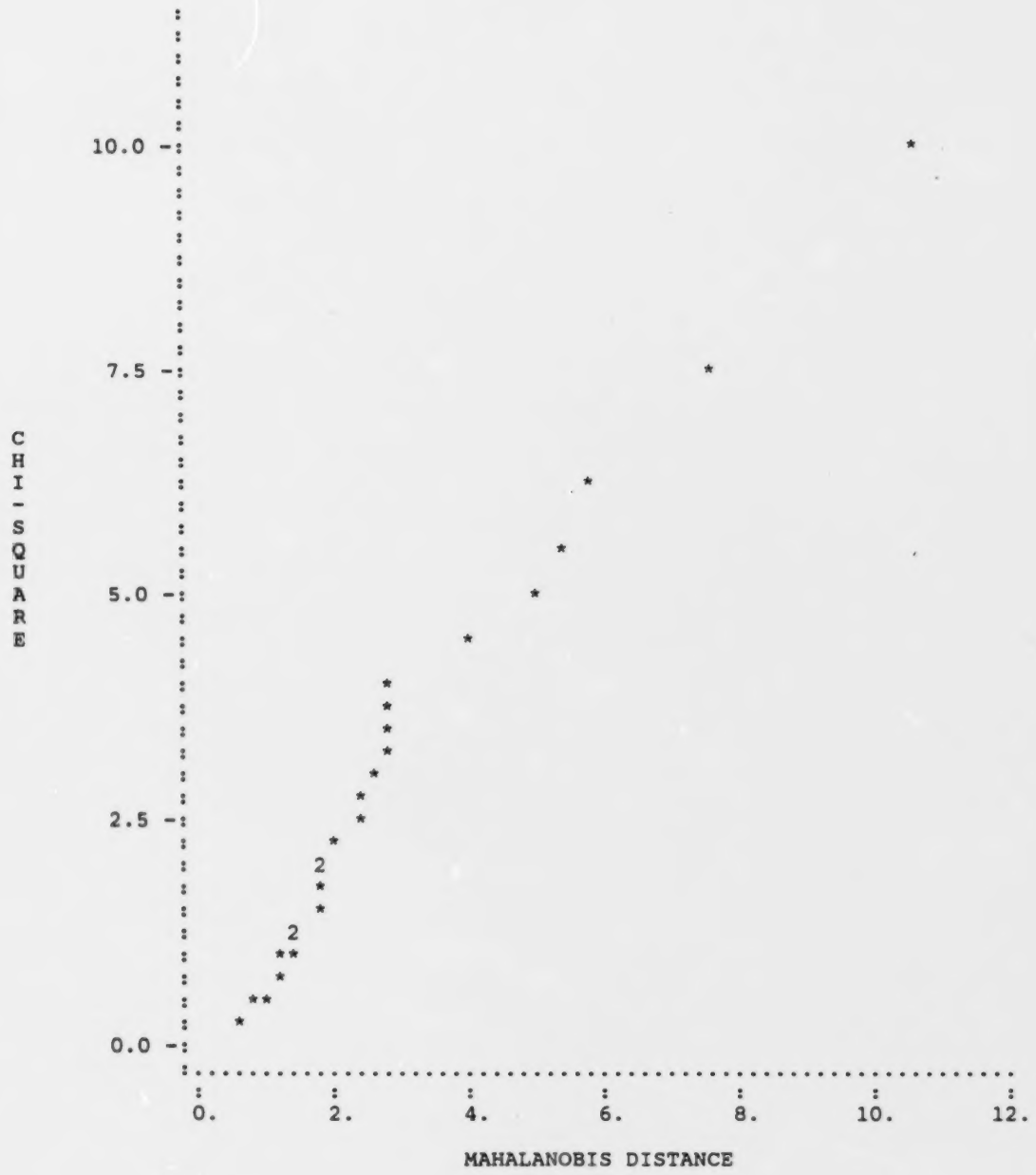
Table 3
Mahalanobis Distances for the 26 Cases

1	5.40434	14	1.28071
2	5.89352	15	2.75819
3	2.67146	16	2.00243
4	1.30686	17	10.53041
5	2.38155	18	3.92584
6	1.79599	19	7.68053
7	2.75163	20	0.82928
8	0.69408	21	1.40629
9	1.19431	22	0.94312
10	4.90097	23	1.42369
11	2.41353	24	2.71666
12	1.77029	25	1.72773
13	2.80861	26	1.78797

Table 4
Sorted D^2 and Associated chi-square and p Values
(df=3 and p percentile = 100(I - .5)/n)

	D Sq	chi sq	p
1	0.69408	0.17988	0.01923
2	0.82928	0.38996	0.05769
3	0.94312	0.56743	0.09615
4	1.19431	0.73313	0.13462
5	1.28071	0.89380	0.17308
6	1.30686	1.05287	0.21154
7	1.40629	1.21253	0.25000
8	1.42369	1.37444	0.28846
9	1.72773	1.53997	0.32692
10	1.77029	1.71044	0.36538
11	1.78797	1.88716	0.40385
12	1.79599	2.07154	0.44231
13	2.00243	2.26515	0.48077
14	2.38155	2.46983	0.51923
15	2.41353	2.68779	0.55769
16	2.67146	2.92176	0.59615
17	2.71666	3.17526	0.63462
18	2.75163	3.45290	0.67308
19	2.75819	3.76095	0.71154
20	2.80861	4.10835	0.75000
21	3.92584	4.50845	0.78846
22	4.90097	4.98259	0.82692
23	5.40434	5.56822	0.86538
24	5.89352	6.34088	0.90385
25	7.68053	7.49482	0.94231
26	10.53041	9.92311	0.98077

Figure 1
Scatterplot of D^2 and chi-square Statistics



APPENDIX A
Output from Program CANSTRAP (Thompson, 1995)

```
aercan.aer
********PROGRAM CANSTRAP                                          04/06/92
        WRITTEN BY BRUCE THOMPSON
                TEXAS A&M UNIVERSITY 77843-4225 AND BAYLOR COLLEGE OF MEDICINE

REFERENCES:
B. THOMPSON. (1984). CANONICAL CORRELATION ANALYSIS. NEWBERRY PARK, CA: SAGE.
B. THOMPSON. (1991). A PRIMER ON THE LOGIC AND USE OF CANONICAL CORRELATION ANALYSIS. MEASUREMENT AND EVALUATION IN COUNSELING
      AND DEVELOPMENT, 24(2), 80-95.
B. THOMPSON. (1995). CANSTRAP: A PROGRAM THAT COMPUTES BOOTSTRAP ESTIMATES OF CANONICAL CORRELATION RESULTS. EDUCATIONAL
      AND PSYCHOLOGICAL MEASUREMENT, 55, 84-94.
B. THOMPSON. (IN PRESS). CANONICAL CORRELATION ANALYSIS: AN EXPLANATION WITH COMMENTS ON CORRECT PRACTICE. IN B. THOMPSON (ED.),
      ADVANCES IN SOCIAL SCIENCE METHODOLOGY. GREENWICH, CT: JAI PRESS.

**JOB TITLE: CANSTRAP RUN USING HOLZINGER & SWINEFORD (1939) DATA
SAMPLE SIZE:            50
N VAR BIGGER SET:        4
N VAR SMALLER SET:       2
OPTIONAL OUTPUT FILE:   99

$$$$$$ INPUT DATA FORMAT= (T14,4F4.0,T8,2F3.0)
  OPTIONAL OUTPUT FORMAT= (8F8.5)

DATA (first and last 5 cases):
  1     115.00000   229.00000     5.00000    24.00000    40.00000     7.00000
  2     126.00000   213.00000     4.00000    20.00000    37.00000     8.00000
  3      93.00000   265.00000    17.00000    18.00000    29.00000     8.00000
  4      91.00000   157.00000     8.00000    16.00000    33.00000     8.00000
  5     114.00000   155.00000     5.00000    24.00000    38.00000     6.00000

 46     101.00000   179.00000     8.00000    24.00000    36.00000     8.00000
 47     103.00000   198.00000    19.00000    26.00000    57.00000    13.00000
 48     140.00000   178.00000    20.00000    29.00000    44.00000     7.00000
 49     119.00000   195.00000    34.00000    24.00000    48.00000    13.00000
 50      85.00000   204.00000    17.00000    25.00000    49.00000     9.00000

CORRELATION MATRIX:
   1 1.00000 0.38474 0.25121 0.37055-0.03992 0.07028
   2 0.38474 1.00000 0.39628 0.28964 0.07621 0.25872
   3 0.25121 0.39628 1.00000 0.54256 0.38925 0.54614
   4 0.37055 0.28964 0.54256 1.00000 0.42966 0.40636
   5-0.03992 0.07621 0.38925 0.42966 1.00000 0.64402
   6 0.07028 0.25872 0.54614 0.40636 0.64402 1.00000
```


BEST COPY AVAILABLE

```
R11 MATRIX AT STEP:   0
  1   1.000  0.385  0.251  0.371
  2   0.385  1.000  0.396  0.290
  3   0.251  0.396  1.000  0.543
  4   0.371  0.290  0.543  1.000

R22 MATRIX AT STEP:   0
  1   1.000  0.644
  2   0.644  1.000

R12 (UPPER RIGHT QUADRANT) MATRIX AT STEP:   0
  1  -0.040  0.070
  2   0.076  0.259
  3   0.389  0.546
  4   0.430  0.406

R21 (LOWER LEFT QUARDRANT) MATRIX AT STEP:   0
  1  -0.040  0.076  0.389  0.430
  2   0.070  0.259  0.546  0.406

R22 INVERTED:
  1   1.709 -1.100
  2  -1.100  1.709

R11 INVERTED:
  1   1.285 -0.401  0.044 -0.384
  2  -0.401  1.323 -0.420 -0.007
  3   0.044 -0.420  1.556 -0.739
  4  -0.384 -0.007 -0.739  1.545

MATRIX TO BE ANALYZED AT STEP:    0
  1   0.170  0.093
  2   0.160  0.273

AT STEP   0
FUNCTION    RCSQ      RC        LAMDA      CHISQ       DF    PCALCULATED

  1        0.35374   0.59476   0.58878     24.10       8    0.00221

  2        0.08894   0.29822   0.91106      4.24       3    0.23689

NOTE: RCSQ EQUALS EIGENVALUE FOR EACH FUNCTION


AT STEP   0
FUNCTION COEFFICIENTS FOR BOTH VARIATES WERE:
  1  -0.318  0.436
  2   0.069  0.507
  3   0.697  0.550
  4   0.479 -0.932
  1   0.366 -1.255
  2   0.724  1.088
```

```
THE DESIGNATED PROCRUSTEAN TARGET MATRIX IS:
  1  -0.318  0.436
  2   0.069  0.507
  3   0.697  0.550
  4   0.479 -0.932
  5   0.366 -1.255
  6   0.724  1.088

AT STEP   0
STRUCTURE COEFFICIENTS FOR BOTH VARIABLE SETS:
  1   0.061  0.424
  2   0.362  0.623
  3   0.905  0.355
  4   0.759 -0.325
  1   0.832 -0.554
  2   0.960  0.280

AT STEP   0
SQUARED STRUCTURE COEFFICIENTS FOR THE FIRST VARIABLE SET:
  1   0.004  0.180
  2   0.131  0.389
  3   0.818  0.126
  4   0.576  0.106

COMMUNALITY COEFFICIENTS:
  1   0.184
  2   0.520
  3   0.944
  4   0.682

REDUNDANCY COEFFICIENTS:
  0   0.135  0.018

POOLED REDUNDANCY COEFFICIENT ACROSS VARIATES:   0.15306

AT STEP   0
SQUARED STRUCTURE COEFFICIENTS FOR THE SECOND VARIABLE SET:

  1   0.693  0.307
  2   0.922  0.078

COMMUNALITY COEFFICIENTS:
  1   1.000
  2   1.000

REDUNDANCY COEFFICIENTS:
  0   0.286  0.017

POOLED REDUNDANCY COEFFICIENT ACROSS VARIATES:   0.30272

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
RESAMPLING #    1
     1    1   18     115.00  186.00   14.00   21.00   22.00    4.00
     1    2   15     123.00  142.00    2.00   15.00   41.00    4.00
     1    3   48     140.00  178.00   20.00   29.00   44.00    7.00
     1    4   41      72.00  121.00    4.00   14.00   32.00    5.00
```

1	5	3	93.00	265.00	17.00	18.00	29.00	8.00
1	6	35	121.00	225.00	23.00	31.00	56.00	13.00
1	7	31	84.00	171.00	31.00	25.00	56.00	14.00
1	8	38	116.00	219.00	6.00	21.00	29.00	8.00
1	9	23	137.00	180.00	10.00	19.00	37.00	8.00
1	10	13	95.00	100.00	1.00	15.00	28.00	5.00
1	11	10	139.00	215.00	12.00	25.00	29.00	6.00
1	12	41	72.00	121.00	4.00	14.00	32.00	5.00
1	13	13	95.00	100.00	1.00	15.00	28.00	5.00
1	14	18	115.00	186.00	14.00	21.00	22.00	4.00
1	15	1	115.00	229.00	5.00	24.00	40.00	7.00
1	16	17	119.00	195.00	6.00	19.00	55.00	13.00
1	17	35	121.00	225.00	23.00	31.00	56.00	13.00
1	18	5	114.00	155.00	5.00	24.00	38.00	6.00
1	19	13	95.00	100.00	1.00	15.00	28.00	5.00
1	20	20	91.00	185.00	23.00	26.00	45.00	12.00
1	21	39	113.00	180.00	37.00	27.00	51.00	9.00
1	22	8	92.00	194.00	23.00	19.00	22.00	5.00
1	23	26	103.00	164.00	16.00	26.00	51.00	10.00
1	24	37	200.00	236.00	30.00	29.00	25.00	7.00
1	25	36	115.00	185.00	25.00	32.00	50.00	14.00
1	26	38	116.00	219.00	6.00	21.00	29.00	8.00
1	27	31	84.00	171.00	31.00	25.00	56.00	14.00
1	28	7	107.00	177.00	26.00	22.00	41.00	11.00
1	29	27	112.00	215.00	9.00	18.00	31.00	8.00
1	30	2	126.00	213.00	4.00	20.00	37.00	8.00
1	31	46	101.00	179.00	8.00	24.00	36.00	8.00
1	32	38	116.00	219.00	6.00	21.00	29.00	8.00
1	33	44	135.00	199.00	28.00	30.00	61.00	10.00
1	34	13	95.00	100.00	1.00	15.00	28.00	5.00
1	35	31	84.00	171.00	31.00	25.00	56.00	14.00
1	36	33	104.00	222.00	20.00	20.00	65.00	10.00
1	37	39	113.00	180.00	37.00	27.00	51.00	9.00
1	38	14	96.00	199.00	18.00	17.00	40.00	11.00
1	39	6	103.00	149.00	9.00	25.00	33.00	8.00
1	40	11	73.00	121.00	9.00	17.00	29.00	5.00
1	41	31	84.00	171.00	31.00	25.00	56.00	14.00
1	42	26	103.00	164.00	16.00	26.00	51.00	10.00
1	43	8	92.00	194.00	23.00	19.00	22.00	5.00
1	44	35	121.00	225.00	23.00	31.00	56.00	13.00
1	45	39	113.00	180.00	37.00	27.00	51.00	9.00
1	46	29	123.00	169.00	10.00	26.00	38.00	9.00
1	47	4	91.00	157.00	8.00	16.00	33.00	8.00
1	48	37	200.00	236.00	30.00	29.00	25.00	7.00
1	49	37	200.00	236.00	30.00	29.00	25.00	7.00
1	50	32	113.00	186.00	24.00	30.00	48.00	11.00

CORRELATION MATRIX:

1	1.00000	0.54443	0.24890	0.49783	-0.15642	-0.08461
2	0.54443	1.00000	0.40560	0.50875	0.11101	0.31844
3	0.24890	0.40560	1.00000	0.69444	0.48024	0.55032
4	0.49783	0.50875	0.69444	1.00000	0.50366	0.57514
5	-0.15642	0.11101	0.48024	0.50366	1.00000	0.81313
6	-0.08461	0.31844	0.55032	0.57514	0.81313	1.00000

R11 MATRIX AT STEP: 1

1	1.000	0.544	0.249	0.498
2	0.544	1.000	0.406	0.509
3	0.249	0.406	1.000	0.694
4	0.498	0.509	0.694	1.000

R22 MATRIX AT STEP: 1

1	1.000	0.813
2	0.813	1.000

R12 (UPPER RIGHT QUADRANT) MATRIX AT STEP: 1

1	-0.156	-0.085
2	0.111	0.318
3	0.480	0.550
4	0.504	0.575

R21 (LOWER LEFT QUARDRANT) MATRIX AT STEP: 1

1	-0.156	0.111	0.480	0.504
2	-0.085	0.318	0.550	0.575

R22 INVERTED:

1	2.951	-2.400
2	-2.400	2.951

R11 INVERTED:

1	1.637	-0.669	0.374	-0.734
2	-0.669	1.633	-0.290	-0.296
3	0.374	-0.290	2.031	-1.449
4	-0.734	-0.296	-1.449	2.522

MATRIX TO BE ANALYZED AT STEP: 1

1	0.220	0.121
2	0.327	0.474

AT STEP 1

FUNCTION	RCSQ	RC	LAMDA	CHISQ	DF	PCALCULATED
1	0.58276	0.76339	0.37071	45.15	8	0.00000
2	0.11152	0.33395	0.88848	5.38	3	0.14599

NOTE: RCSQ EQUALS EIGENVALUE FOR EACH FUNCTION

AT STEP 1
FUNCTION COEFFICIENTS FOR BOTH VARIATES WERE:

1	-0.755	0.041
2	0.251	1.144
3	0.216	-0.052
4	0.856	-0.489
1	0.259	-1.698
2	0.778	1.532

```
SAMPLE#    1 FUNCTION MATRIX AFTER ROTATION:
  1  -0.747  0.118
  2   0.365  1.112
  3   0.210 -0.074
  4   0.801 -0.574
  1   0.085 -1.716
  2   0.929  1.445

AT STEP   1
STRUCTURE COEFFICIENTS FOR BOTH VARIABLE SETS:
  1  -0.097  0.419
  2   0.451  0.855
  3   0.728  0.008
  4   0.761 -0.001
  1   0.841 -0.541
  2   0.999  0.050

AT STEP   1
SQUARED STRUCTURE COEFFICIENTS FOR THE FIRST VARIABLE SET:
  1   0.009  0.176
  2   0.204  0.730
  3   0.530  0.000
  4   0.579  0.000

COMMUNALITY COEFFICIENTS:
  1   0.185
  2   0.934
  3   0.531
  4   0.579

REDUNDANCY COEFFICIENTS:
  1   0.193  0.025

POOLED REDUNDANCY COEFFICIENT ACROSS VARIATES:   0.21795

AT STEP   1
SQUARED STRUCTURE COEFFICIENTS FOR THE SECOND VARIABLE SET:

  1   0.707  0.293
  2   0.998  0.002

COMMUNALITY COEFFICIENTS:
  1   1.000
  2   1.000

REDUNDANCY COEFFICIENTS:
  1   0.497  0.016

POOLED REDUNDANCY COEFFICIENT ACROSS VARIATES:   0.51322

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
RESAMPLING #    2
     2    1   13      95.00  100.00    1.00   15.00   28.00    5.00
     2    2   36     115.00  185.00   25.00   32.00   50.00   14.00
     2    3   25     156.00  252.00   36.00   26.00   45.00   13.00
     2    4   44     135.00  199.00   28.00   30.00   61.00   10.00
```

2	5	11	73.00	121.00	9.00	17.00	29.00	5.00
2	6	21	119.00	240.00	24.00	23.00	39.00	11.00
2	7	49	119.00	195.00	34.00	24.00	48.00	13.00
2	8	23	137.00	180.00	10.00	19.00	37.00	8.00
2	9	19	147.00	207.00	23.00	30.00	26.00	12.00
2	10	3	93.00	265.00	17.00	18.00	29.00	8.00
2	11	9	117.00	310.00	18.00	27.00	31.00	6.00
2	12	33	104.00	222.00	20.00	20.00	65.00	10.00
2	13	9	117.00	310.00	18.00	27.00	31.00	6.00
2	14	24	139.00	204.00	14.00	18.00	38.00	13.00
2	15	28	98.00	139.00	18.00	27.00	44.00	11.00
2	16	37	200.00	236.00	30.00	29.00	25.00	7.00
2	17	33	104.00	222.00	20.00	20.00	65.00	10.00
2	18	22	117.00	152.00	17.00	20.00	48.00	10.00
2	19	7	107.00	177.00	26.00	22.00	41.00	11.00
2	20	27	112.00	215.00	9.00	18.00	31.00	8.00
2	21	18	115.00	186.00	14.00	21.00	22.00	4.00
2	22	48	140.00	178.00	20.00	29.00	44.00	7.00
2	23	12	87.00	203.00	20.00	22.00	44.00	10.00
2	24	49	119.00	195.00	34.00	24.00	48.00	13.00
2	25	25	156.00	252.00	36.00	26.00	45.00	13.00
2	26	36	115.00	185.00	25.00	32.00	50.00	14.00
2	27	9	117.00	310.00	18.00	27.00	31.00	6.00
2	28	27	112.00	215.00	9.00	18.00	31.00	8.00
2	29	22	117.00	152.00	17.00	20.00	48.00	10.00
2	30	10	139.00	215.00	12.00	25.00	29.00	6.00
2	31	26	103.00	164.00	16.00	26.00	51.00	10.00
2	32	12	87.00	203.00	20.00	22.00	44.00	10.00
2	33	38	116.00	219.00	6.00	21.00	29.00	8.00
2	34	16	108.00	227.00	27.00	30.00	54.00	11.00
2	35	25	156.00	252.00	36.00	26.00	45.00	13.00
2	36	10	139.00	215.00	12.00	25.00	29.00	6.00
2	37	15	123.00	142.00	2.00	15.00	41.00	4.00
2	38	15	123.00	142.00	2.00	15.00	41.00	4.00
2	39	10	139.00	215.00	12.00	25.00	29.00	6.00
2	40	27	112.00	215.00	9.00	18.00	31.00	8.00
2	41	3	93.00	265.00	17.00	18.00	29.00	8.00
2	42	47	103.00	198.00	19.00	26.00	57.00	13.00
2	43	34	110.00	161.00	16.00	33.00	49.00	8.00
2	44	45	110.00	199.00	6.00	26.00	51.00	12.00
2	45	15	123.00	142.00	2.00	15.00	41.00	4.00
2	46	33	104.00	222.00	20.00	20.00	65.00	10.00
2	47	23	137.00	180.00	10.00	19.00	37.00	8.00
2	48	1	115.00	229.00	5.00	24.00	40.00	7.00
2	49	34	110.00	161.00	16.00	33.00	49.00	8.00
2	50	45	110.00	199.00	6.00	26.00	51.00	12.00

CORRELATION MATRIX:

1	1.00000	0.23345	0.33778	0.28308	-0.19063	0.06078
2	0.23345	1.00000	0.37939	0.24550	-0.16102	0.11065
3	0.33778	0.37939	1.00000	0.53354	0.32025	0.65803
4	0.28308	0.24550	0.53354	1.00000	0.24046	0.41137
5	-0.19063	-0.16102	0.32025	0.24046	1.00000	0.55515
6	0.06078	0.11065	0.65803	0.41137	0.55515	1.00000

R11 MATRIX AT STEP: 2

1	1.000	0.233	0.338	0.283
2	0.233	1.000	0.379	0.246
3	0.338	0.379	1.000	0.534
4	0.283	0.246	0.534	1.000

R22 MATRIX AT STEP: 2

1	1.000	0.555
2	0.555	1.000

R12 (UPPER RIGHT QUADRANT) MATRIX AT STEP: 2

1	-0.191	0.061
2	-0.161	0.111
3	0.320	0.658
4	0.240	0.411

R21 (LOWER LEFT QUARDRANT) MATRIX AT STEP: 2

1	-0.191	-0.161	0.320	0.240
2	0.061	0.111	0.658	0.411

R22 INVERTED:

1	1.445	-0.802
2	-0.802	1.445

R11 INVERTED:

1	1.163	-0.135	-0.257	-0.159
2	-0.135	1.187	-0.377	-0.052
3	-0.257	-0.377	1.596	-0.686
4	-0.159	-0.052	-0.686	1.424

MATRIX TO BE ANALYZED AT STEP: 2

1	0.173	0.061
2	0.218	0.456

AT STEP 2

FUNCTION	RCSQ	RC	LAMDA	CHISQ	DF	PCALCULATED
1	0.49653	0.70465	0.43691	37.68	8	0.00001
2	0.13222	0.36362	0.86778	6.45	3	0.09156

NOTE: RCSQ EQUALS EIGENVALUE FOR EACH FUNCTION

AT STEP 2
FUNCTION COEFFICIENTS FOR BOTH VARIATES WERE:

1	-0.304	0.656
2	-0.258	0.655
3	1.013	0.073
4	0.190	-0.267
1	0.169	-1.190
2	0.897	0.801

```
SAMPLE#    2 FUNCTION MATRIX AFTER ROTATION:
  1  -0.427  0.583
  2  -0.382  0.592
  3   0.979  0.271
  4   0.238 -0.224
  1   0.400 -1.134
  2   0.721  0.962

AT STEP   2
STRUCTURE COEFFICIENTS FOR BOTH VARIABLE SETS:
  1  -0.118  0.749
  2  -0.052  0.776
  3   0.817  0.573
  4   0.546  0.231
  1   0.800 -0.600
  2   0.943  0.332

AT STEP   2
SQUARED STRUCTURE COEFFICIENTS FOR THE FIRST VARIABLE SET:
  1   0.014  0.562
  2   0.003  0.602
  3   0.667  0.329
  4   0.298  0.053

COMMUNALITY COEFFICIENTS:
  1   0.576
  2   0.605
  3   0.996
  4   0.352

REDUNDANCY COEFFICIENTS:
  2   0.122  0.051

POOLED REDUNDANCY COEFFICIENT ACROSS VARIATES:   0.17302

AT STEP   2
SQUARED STRUCTURE COEFFICIENTS FOR THE SECOND VARIABLE SET:

  1   0.640  0.360
  2   0.889  0.111

COMMUNALITY COEFFICIENTS:
  1   1.000
  2   1.000

REDUNDANCY COEFFICIENTS:
  2   0.380  0.031

POOLED REDUNDANCY COEFFICIENT ACROSS VARIATES:   0.41084

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
RESAMPLING #    3

* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
RESAMPLING #    4
```

BEST COPY AVAILABLE

```
* * * *
```

```
     SUBJECT #  25   1028 TIMES
     SUBJECT #  26    973 TIMES
     SUBJECT #  27    976 TIMES
     SUBJECT #  28   1005 TIMES
     SUBJECT #  29    993 TIMES
     SUBJECT #  30    980 TIMES
     SUBJECT #  31    984 TIMES
     SUBJECT #  32    998 TIMES
     SUBJECT #  33   1008 TIMES
     SUBJECT #  34   1006 TIMES
     SUBJECT #  35   1019 TIMES
     SUBJECT #  36   1034 TIMES
     SUBJECT #  37    942 TIMES
     SUBJECT #  38    971 TIMES
     SUBJECT #  39    986 TIMES
     SUBJECT #  40    994 TIMES
     SUBJECT #  41    991 TIMES
     SUBJECT #  42   1023 TIMES
     SUBJECT #  43    975 TIMES
     SUBJECT #  44    958 TIMES
     SUBJECT #  45   1028 TIMES
     SUBJECT #  46   1010 TIMES
     SUBJECT #  47   1020 TIMES
     SUBJECT #  48   1019 TIMES
     SUBJECT #  49   1023 TIMES
     SUBJECT #  50    988 TIMES

THE MEAN USAGE OF SUBJECTS WAS   999.9971
THE SD OF USAGE WAS    26.2676
THE MINIMUM N OF TIMES A SUBJECT WAS RESAMPLED WAS   942.0000
THE MAXIMUM N OF TIMES A SUBJECT WAS RESAMPLED WAS  1056.0000


*** SUMMARY STATISTICS FOR Rc SQUARED:
   *** MEANs
      0.43958  0.13158
   *** SDs
      0.11744  0.07873
   *** SKEWNESSs
      0.23703  0.85067
   *** KURTOSISs
     -0.09252  0.60673

LARGER VARIABLE SET****
FUNCTION COEFFICIENTS:
   *** MEANs
   1 -0.2394  0.3141
   2 -0.0147  0.3504
   3  0.3953  0.4304
   4  0.4196 -0.6641
   *** SDs
   1  0.3332  0.4803
   2  0.2755  0.3996
   3  0.5256  0.3948
   4  0.3677  0.4173
```

BEST COPY AVAILABLE

```
   *** SKEWNESSs
   1  0.1812 -0.3334
   2  0.1025 -0.6003
   3 -1.2935 -0.7817
   4 -0.6616  1.1365
   *** KURTOSISs
   1 -0.4745 -0.3368
   2 -0.0576  0.2432
   3  0.7095  0.9570
   4  0.5870  1.4635


*** SUMMARY STATISTICS FOR STRUCTURE COEFFICIENTS:
   *** MEANs
   1  0.0146  0.2913
   2  0.1574  0.4520
   3  0.5515  0.3099
   4  0.5407 -0.2115
   *** SDs
   1  0.2872  0.4174
   2  0.3872  0.2901
   3  0.6003  0.2962
   4  0.4675  0.3532
   *** SKEWNESSs
   1 -0.1697 -0.5314
   2 -0.6357 -0.7588
   3 -1.7525 -0.0483
   4 -1.9404  0.4557
   *** KURTOSISs
   1 -0.1597 -0.2460
   2 -0.3010  0.6731
   3  1.5082 -0.1744
   4  2.4797 -0.2703

SMALLER VARIABLE SET****
FUNCTION COEFFICIENTS:
   *** MEANs
   1  0.4698 -1.1654
   2  0.3082  1.0684
   *** SDs
   1  0.3643  0.3221
   2  0.7101  0.2740
   *** SKEWNESSs
   1  1.0118  3.0347
   2 -1.2058 -0.4001
   *** KURTOSISs
   1  0.5090 12.8658
   2  0.0569  0.3770


*** SUMMARY STATISTICS FOR STRUCTURE COEFFICIENTS:
   *** MEANs
   1  0.6552 -0.4587
   2  0.6218  0.3074
```

BEST COPY AVAILABLE

```
*** SDs
1  0.4766  0.3654
2  0.6555  0.2994
*** SKEWNESSs
1 -1.9942  1.5918
2 -1.8525  0.1646
*** KURTOSISs
1  2.4850  2.8181
2  1.7368  0.1547
```

47

48